Introduction to the Issue on Spatial Audio

SPATIAL audio is an area that has gained in popularity in the recent years. Audio reproduction setups have evolved from the traditional two-channel stereophonic setup towards multi-channel loudspeaker setups. Advances in acoustic signal processing even made it possible to create surround sound listening experiences as well as attempts to create binaural real life listening experiences using traditional stereo speakers and headphones. Finally, there has been an increased interest in creating different sound zones in the same acoustic space. At the same time, the computational capacity provided by mobile audio playback devices has increased significantly. These developments enable new possibilities for advanced audio signal processing, such that in the future we can record, transmit and reproduce spatial audio in ways that have not been possible before. In addition, there have been fundamental advances in our understanding of 3D audio.

Due to the increasing number of different formats and reproduction systems for spatial audio, ranging from headphones to 22.2 speakers, it is a major challenge to ensure interoperability between formats and systems, and consistent delivery of high quality spatial audio. Therefore, the MPEG committee has established a new standard for 3D audio coding. The first paper by Herre *et al.* in this special issue is an invited paper that introduces the MPEG-H 3D Audio standard and the underlying principles.

Some of the fundamental problems in spatial audio are associated with scene analysis—the capture of a scene for later reproduction, finding the location of a sound source, isolating the sound of that source. Microphones, and in particular microphone arrays are widely used to capture and to decompose spatial sound scenes into their components. In this sense they perform a role similar to that cameras do in the analysis of visual scenes. They thus form a valuable tool in the study and application of spatial audio. These arrays can be used for capturing spatial sound in terms of spatial basis functions of the underlying basis functions of the wave equation, without installing many microphones over the wide area. In particular, an open spherical microphone array, in which microphones are installed over the surface of a sphere, can capture the sound from every direction at the same time. However, the algorithms used to analyze the signals via beamforming have numerical instabilities because of the roots of spherical Bessel functions. Chardon *et al.* develop a design method for the array that adds some microphones inside the sphere to alleviate the instability. A crucial issue in designing a microphone array is the knowledge of the relative coordinates of microphones that are installed in a microphone array. The problem of determining these coordinates is termed array calibration, and has to be performed either manually, or via the solution of geometric problems that for large numbers of sources and receivers cannot be solved in a closed form but must be tackled via nonlinear constrained optimization. These routines are subject to being trapped in local minima. Simayijiang *et al.* provide a closed form solution for various cases of an array that is constructed of dual microphones that are available as a pair (e.g., as in a mobile phone with two microphones installed at a known distance apart).

A particular problem in acoustic beamforming and source localization is the deleterious influence of reverberation on the performance of many algorithms. This problem is sought to be addressed by several authors in this issue. A first group of papers exploit prior geometric knowledge about the enclosure. Taghizadeh *et al.* propose a method to localize a single sound source using synchronous and asynchronous recordings of a single microphone in a reverberant enclosure. In this work the multipath propagation is modelled using multiple virtual microphones (as images of the real microphone). A multipath distance matrix is then constructed whose elements consist of the squared distances between the pairs of microphones (real or virtual) or the squared distances between the microphones and the source. The distances between the actual and virtual microphones are computed from the geometry of the enclosure and the position of the microphone. Source localization is achieved through optimizing the location of the source matching those measurements. The narrowband localization of sound sources in reverberant environments is a challenging task in the presence of reverberation. In previous works it has been shown that prior knowledge about the room can be exploited by dictionary-based localization methods. Chardon *et al.* show that these methods fail to accurately localize a source when the measurements are done at frequencies close to modal frequencies of the room. The authors propose a new model for the acoustic sound field in case both the room geometry and boundary conditions are unknown using the Vekua theory that allows to decompose the sound field in a direct and reverberant sound field. Based on this model a dereverberation pre-processing step is developed that allows to remove the reverberation from acoustical measurements without any prior on the room or on the signals. This pre-processing step is compatible with various localization methods, and enables the narrowband source localization in a unknown reverberant environment.

The paper of Dokmanic *et al.* assumes that some methods for locating early reflections for the source of interest are available, and then seeks to develop various beamforming algorithms that can take advantage of this knowledge. Using an inexact analogy with ideas of rake receivers in wireless communication, they term their beamformers as an "acoustic rake receivers". They then extend several standard beamforming frameworks such as delay-and-sum, minimum variance distortionless response, and Signal-to-Interference-and-Noise Ratio in various ways to incorporate information about the reflections. While

Digital Object Identifier 10.1109/JSTSP.2015.2447112

their formulation is general, simulations are presented for two-dimensional geometries, and show impressive gains. Most beamforming algorithms use noise models that are idealized and Gaussian. On the other hand the noise spectrum encountered in practice shows significant departure from Gaussianity. In practical conditions, the noise may be diffuse or arrive from a point source, and is further confounded by late reverberation. In order to improve the reliability of direction-of-arrival (DOA) estimation in various adverse noisy conditions, Xue *et al.* propose a novel DOA estimation method in this paper, via the use of a "Weighted Bispectrum Spatial Correlation Matrix (WBSCM)," which contains higher-order statistics (HOS) information of the signal. The WBSCM reflects the spatial correlation of the bispectrum phase differences between different microphones. This statistic deals with both Gaussian and non-Gaussian noise, as the HOS of the Gaussian signal is theoretically zero. The paper develops various algorithms to use this statistic for the processing of speech signals in the presence of various kinds of noise, and shows impressive results.

Besides the traditional techniques for spatial sound reproduction, like stereophony or Sound Field Synthesis, parametric techniques aim at the dynamic decomposition of a captured sound field into spatio-temporal components. These components are often inspired by human spatial perception. A well known representative of these methods is Directional Audio Coding. Politis *et al.* present an extension of the method which provides a higher separation between simultaneous sources and reverberation. It is based on a sectorial analysis of the sound field using spherical harmonics signals which may have been captured by an spherical microphone array. The results of a listening experiment prove the performance of the proposed analysis scheme. In the loudspeaker reproduction, the reverberation of the reproduction room sometimes gives detrimental effects on the reproduced sound. Grosse *et al.* developed a rendering method of both source and room acoustics, which minimizes the difference of binaurally recorded sounds in the recorded and reproduction rooms through auditory filter banks. To control the reverberant sound in the reproduction, they use the dipole loudspeaker whose null is set to the listener's direction. In sound field recording and reproduction we sometimes have prior information about the location of the sound sources. Koyama *et al.* propose a recording and rendering method that exploits such prior information. The proposed method for planar and linear arrays of microphones and loudspeakers optimizes spatial basis functions and their coefficients, which represent the driving signals of loudspeakers, on the basis of the maximum a posteriori estimation approach, leading to higher reproduction accuracy above the spatial Nyquist frequency when there are fewer microphones than loudspeakers.

In the spatial audio reproduction, many loudspeakers are often required to reproduce sound field accurately, which may be possible only in the theaters or specially designed laboratories. An alternative is a binaural audio display that attempts to recreate the sound pressure at the listener's eardrum. The Head-Related Impulse Responses (HRIRs) or Binaural Room Impulse Response (BRIRs) are usually used as fundamental data in binaural audio rendering. The Head-Related Transfer Function (HRTF) is a frequency-domain representation of the HRIR. Among papers addressing this technology, Andreopoulou *et al.* investigate the variation of measured HRTF data among 10 laboratories using the same dummy head and find the magnitude and timing differences between left and right ear data caused by the difference in measurement technique equipment and post-processing method. On the other hand, MPEG has launched MPEG-H 3D Audio to standardize coding and rendering methods for spatial audio. As a part of this standardization, Lee *et al.* develop a binaural rendering method for mobile audio applications, which renders the multichannel audio signal, such as 22.2 sound, into the binaural signal with BRIR. Identifying the relevant features in HRTFs is an ongoing topic of research for decades. Romigh *et al.* investigate the effect of spatial smoothing on localization properties of measured HRTFs. Smoothing is performed by truncating the series expansion with respect to spherical harmonics. The reported localization experiment indicates that accurate localization can still be performed even when severe truncation and hence smoothing takes place. These results suggest that the fine structure of HRTFs is not too important and may be omitted for an efficient representation. It is typical to represent the HRTF's for a listener facing forwards. Brinkmann *et al.* show that head rotation of the listener affects the HRTFs, and it should be taken into account in binaural rendering. In addition, they present a technique how this can be achieved in interactive acoustic simulations. The final paper in this Spatial Audio Special Issue is by Romigh *et al.* and it deals with high-quality real-time HRTF-based sound reproduction. The focus is on localization accuracy, and the results show that it is possible to achieve similar performance with headphones as is possible in real free-field listening.

We would like to thank all the peer-reviewers for their diligent work and the helpful staff at IEEE JSTSP. These contributions to this Special Issue have been invaluable.

LAURI SAVIOJA, *Lead Guest Editor*
Department of Computer Science
Aalto University
02150 Espoo, Finland

AKIO ANDO, *Guest Editor*
University of Toyama
Toyama 930-0887, Japan

RAMANI DURAISWAMI, *Guest Editor*
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742 USA

EMANUËL A. P. HABETS, *Guest Editor*
International Audio Laboratories Erlangen
91058 Erlangen, Germany

SASCHA SPORS, *Guest Editor*
Institute of Telecommunications Engineering
Universität Rostock
18051 Rostock, Germany

Lauri Savioja (M'00–SM'08) received the degrees of M.Sc., the Licentiate of Science, and the Doctor of Science in Technology, from the Helsinki University of Technology (TKK), Espoo, Finland, in 1991, 1995, and 1999, respectively. In all those degrees, he majored in computer science. The topic of his doctoral thesis was room acoustic modeling. He worked at the TKK Laboratory of Telecommunications Software and Multimedia as a Researcher, Lecturer, and Professor from 1995 till the formation of the Aalto University where he currently works as a Professor in the Department of Computer Science, School of Science. The academic year 2009–2010, he spent as a visiting researcher at NVIDIA Research. His research interests include room acoustics, virtual reality, and parallel computation. Prof. Savioja is a fellow of the Audio Engineering Society (AES), and a life member of the Acoustical Society of Finland. From 2010 to 2013 he acted as an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.

Akio Ando (M'80) received his B.S. and M.S. degrees from Kyushu Institute of Design, Japan, in 1978 and 1980, respectively. He also received a Dr. Eng. degree from Toyohashi University of Technology, Japan, in 2001. In 1980, he joined the Japan Broadcasting Corporation (NHK) and worked at NHK Science and Technology Research Laboratories from 1983 to 2013. He supervised the development of a 22.2 multichannel sound system for an ultra-high definition TV. Since 2013, he has been a Professor at University of Toyama, Japan. His research interests include signal processing, electroacoustical transducers and spatial sound reproduction. Prof. Ando was the vice president of the Acoustical Society of Japan (ASJ) from 2013 to 2015.

Ramani Duraiswami (M'99) received his undergraduate degree in 1985 from the Indian Institute of Technology, Bombay, and his Ph.D. degree in 1991 from The Johns Hopkins University. He is a Professor of Computer Science and in the Institute for Advanced Computer Studies at the University of Maryland, College Park. He is also the Director of the Perceptual Interfaces and Reality Laboratory, at the University, which he established in 2001. Prof. Duraiswami has wide scientific interests in scientific computation (fast multipole methods, GPU and heterogeneous computing); in the simulation and understanding of the three dimensional auditory perception of acoustic scenes (HRTF personalization via measurement, machine learning and simulation; room impulse response modeling; the development of efficient engines for creating virtual auditory spaces); in the interplay of the visual and auditory modalities in joint scene analysis; and in the development of new devices and instrumentation. Some of his research results have been spun out commercially.

Emanuël A. P. Habets (S'02–M'07–SM'11) received the B.Sc. degree in electrical engineering from the Hogeschool Limburg, The Netherlands, in 1999, and the M.Sc. and Ph.D. degrees in electrical engineering from the Technische Universiteit Eindhoven (TU/e), The Netherlands, in 2002 and 2007, respectively. From 2007 until 2009, he was a Postdoctoral Fellow at the Technion-Israel Institute of Technology and at the Bar-Ilan University, Israel. From 2009 until 2010, he was a Research Fellow at Imperial College London, U.K. Currently, he is an Associate Professor at the International Audio Laboratories Erlangen (a joint institution of Fraunhofer IIS and the University of Erlangen-Nuremberg, Germany) and Head of the Spatial Audio Research Group at Fraunhofer IIS. He holds several patents, and has authored or coauthored several book chapters and more than 150 papers in journals and conference proceedings. His research activities center around audio and acoustic signal processing, and include spatial sound recording and reproduction, speech enhancement (dereverberation, noise reduction, echo reduction), and sound localization and tracking. Dr. Habets was a member of the organization committee of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC) in Eindhoven, The Netherlands, a general co-chair of the 2013 International Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in New Paltz, New York, general co-chair of the 2014 International Conference on Spatial Audio (ICSA) in Erlangen, Germany, and a co-organizer of the 2014 REVERB Challenge. He is a Senior Member of the IEEE, a member of the Audio Engineering Society (AES), a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2011–2016) and a member of the IEEE Signal Processing Society Standing Committee on Industry Digital Signal Processing Technology (2013–2015). Currently, he serves as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. He is the recipient, with I. Cohen and S. Gannot, of the 2014 IEEE SPS Signal Processing Letters Best Paper Award.

Sascha Spors received the Dipl.-Ing. degree in electrical engineering and the Dr.-Ing. degree with distinction from the University of Erlangen-Nuremberg, Erlangen, Germany, in 2000 and 2006, respectively. Currently, he heads the virtual acoustics and signal processing group as a full Professor at the Institute of Telecommunications Engineering, Universität Rostock, Rostock, Germany. From 2005 to 2012, he was heading the audio technology group as a Senior Research Scientist at the Telekom Innovation Laboratories, Technische Universität Berlin, Berlin, Germany. From 2001 to 2005, he was a member of the research staff at the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg. He holds several patents, and has authored or coauthored several book chapters and more than 200 papers in journals and conference proceedings. His current areas of interest include sound field analysis and reproduction using multichannel techniques, the perception of synthetic sound fields, and efficient multichannel digital signal processing. Prof. Spors is a member of the Audio Engineering Society (AES) and the German Acoustical Society (DEGA). He was awarded the Lothar Cremer prize of the German Acoustical Society in 2011 and the Board of Governors award of the AES in 2015. Prof. Spors is Co-chair of the AES Technical Committee on Spatial Audio and an Associate Technical Editor of the *Journal of the Audio Engineering Society* and the IEEE SIGNAL PROCESSING LETTERS.